Non-Stochastic Hypothesis Testing with Application to Privacy Against Hypothesis-Testing Adversaries

Farhad Farokhi

***Abstract*— We consider privacy against hypothesis-testing adversaries within a non-stochastic framework. We develop a theory of non-stochastic hypothesis testing by borrowing the notion of uncertain variables from non-stochastic information theory. We define tests as binary-valued mappings on uncertain variables and prove a fundamental bound on the performance of tests in non-stochastic hypothesis testing. We use this bound to develop a measure of privacy. We then construct reporting policies with prescribed privacy and utility guarantees. The utility of a reporting policy is measured by the distance between reported and original values. We illustrate the effects of using such privacy-preserving reporting polices on a publicly-available practical dataset of preferences and demographics of young individuals with Slovakian nationality.**

I. Introduction

For decades, stochastic policies have been used for privacy protection [1]. More recently, stochastic policies with provable privacy guarantees have been developed using differential privacy [2]–[6] and information-theoretic privacy [7]–[15]. Differential privacy uses randomization to ensure that the statistics of reported outputs do not change noticeably by variations in an individual entry of the dataset. This is often ensured by the use of additive Laplace or Gaussian noise with a scale proportional to the sensitivity of reports on a private dataset with respect to an individual entry. Information-theoretic privacy, dating back to the secrecy problem [16], emphasizes on masking or equivocating of information from the intended primary receiver or a secondary receiver with as much information as the primary receiver (e.g., an eavesdropper) while providing guarantees on utility by bounding distortion, i.e., the distance between obfuscated and original reports [14], [17]–[19] .

Although the above-mentioned stochastic policies provide provable privacy guarantees, many organizations still use deterministic heuristic-based privacy-preserving methods, such as k-anonymity [20], [21] and ℓ-diversity [22]. For instance, anonymization is frequently used by governments[1] or companies[2] alike for releasing private data to the broader public for analysis even though it is proved to be insufficient for privacy preservation [23]–[25]. Other policies, such as k-anonymity, are also vulnerable to attacks [22], [26].

F. Farokhi is with the Data61 at the Commonwealth Scientific and Industrial Research Organisation (CSIRO) and the Department of Electrical and Electronic Engineering at the University of Melbourne, Australia

e-mails: farhad.farokhi@data61.csiro.au; farhad.farokhi@unimelb.edu.au

[1]See `https://data.gov.au` and `https://www.data.gov` for examples of government initiative.

[2]See `https://www.kaggle.com` for examples of data from private companies and individuals.

The popularity of non-stochastic/deterministic privacy-preserving policies is perhaps caused by factors, such as undesirable properties of differentially-private additive noise especially Laplace noise [27], [28], simplicity of implementing deterministic policies (in the sense of not requiring expertise in probability theory) [29], and generation of unreasonable/unrealistic outputs by the use of randomness [30]–[32]. The guarantees of information-theoretic privacy are also most often presented in the form of averages, i.e., bounds on the average amount of leaked information. Information-theoretic privacy also requires the knowledge of the probability distribution of private dataset, which might not be available at the time of design or might change over time. These observations motivate the need for better understanding of non-stochastic privacy policies.

Absence of systematic methods for developing or assessing deterministic privacy-preserving policies is due to the lack of privacy measures for deterministic policies on non-stochastic arbitrary datasets. This makes proving privacy guarantees for deterministic policies, in any sense, even if weak or limited in scope or practice, impossible. Recently, non-stochastic information theory (see, e.g., [33]–[40]) was used to develop a deterministic measure of privacy in [26]. This measure was successfully utilized to show that binning, a popular deterministic policy for privacy preservation, provides some guarantees, and to prove that k-anonymity is in fact *not* privacy preserving, without resorting to extensive simulations and numerical studies (which are only sufficient and not necessary in the analysis of general policies). The privacy measure in [26] is perfect for providing protections against generic adversaries; however, in some instances, more might be known about the privacy-intrusive adversaries, hence the privacy measure can be further refined.

A category of adversaries studied in privacy literature is the category of hypothesis testing adversaries [4], [12], [41]. Such an adversary is interested in examining the validity of a hypothesis, e.g., if a house is occupied or if an individual has a certain disease based on some reports. The privacy risk, in this case, is often measured by the minimum error probability of the adversary. In this paper, we expand this analysis to a non-stochastic framework. To do so, we develop a theory of non-stochastic hypothesis testing by borrowing the concept of uncertain variables from non-stochastic information theory. Uncertain variables only consider support sets and do not assign distributions/measures to variables. In non-stochastic hypothesis theory, we define tests as binary-valued functions on uncertain variables. We prove a fundamental bound for the performance of tests. This bound is used to develop a

measure of privacy. We then construct reporting functions with given privacy and utility guarantees. We illustrate the outcomes of using such privacy-preserving polices on practical datasets.

The rest of the paper organized as follows. We present the non-stochastic hypothesis testing framework in Section II. In Section III, we investigate privacy against hypothesis-testing adversaries. Finally, we present numerical results in Section IV and conclude the paper in Section V.

II. Non-Stochastic Hypothesis Testing

In this section, we develop a framework for non-stochastic hypothesis testing starting by introducing the notion of uncertain variables.

A. Uncertain Variables

Consider sample space Ω whose elements $\omega \in \Omega$ are the source of uncertainty. An uncertain variable is a mapping defined over Ω, such as $X : \Omega \to \mathbb{X}$, with $X(\omega)$ denoting a realization of the uncertain variable. When the dependence of the uncertain variable, or in short u.v., to the sample is evident from the context, $X(\omega)$ is replaced by X. In this paper, we restrict ourselves to real-valued uncertain variables, e.g., $\mathbb{X} \subseteq \mathbb{R}^{n_x}$ for some integer $n_x \geq 0$. *Marginal range* of any uncertain variable X is $[\![X]\!] := \{X(\omega) : \omega \in \Omega\} \subseteq \mathbb{X}$. *Joint range* of any two uncertain variables $X : \Omega \to \mathbb{X}$ and $Y : \Omega \to \mathbb{Y}$ is $[\![X, Y]\!] := \{(X(\omega), Y(\omega)) : \omega \in \Omega\} \subseteq \mathbb{X} \times \mathbb{Y}$. *Conditional range* of any uncertain variable X, conditioned on the realization of another uncertain variable $Y(\omega) = y$, is $[\![X|y]\!] := \{X(\omega) : \exists \omega \in \Omega \text{ such that } Y(\omega) = y\} \subseteq [\![X]\!]$. Uncertain variables $(X_i)_{i=1}^n$ are *unrelated* if $[\![X_1, \ldots, X_n]\!] = [\![X_1]\!] \times \cdots \times [\![X_n]\!]$. Further, they are conditionally unrelated, conditioned on uncertain variable Y, if $[\![X_1, \ldots, X_n|y]\!] = [\![X_1|y]\!] \times \cdots \times [\![X_n|y]\!]$ for all $y \in [\![Y]\!]$. For two uncertain variables, X_1 and X_2 are unrelated if $[\![X_1|x_2]\!] = [\![X_1]\!], \forall x_2 \in [\![X_2]\!]$.

An uncertain variable X for which $[\![X]\!]$ is uncountably infinite is a *continuous* uncertain variable, similar to a continuous random variable. An uncertain variable X for which $[\![X]\!]$ is finite is a *discrete* uncertain variable. Non-stochastic entropy of a continuous uncertain variable X can be defined as

$$h_0(X) := \log_e(\mu([\![X]\!])) \in \mathbb{R} \cup \{\pm\infty\}, \tag{1}$$

where μ is the Lebesgue measure. While the logarithm can be taken in any basis, it is in the natural basis in this paper following the literature on differential entropy of continuous random variables. The non-stochastic entropy in (1) is sometimes referred to as Rényi differential 0-entropy [36]. Non-stochastic entropy of a discrete uncertain variable X can be defined as

$$H_0(X) := \log_2(|[\![X]\!]|) \in \mathbb{R}, \tag{2}$$

where $|\cdot|$ is the cardinality of a set. In this paper, for discrete uncertain variables, in line with the literature on entropy of discrete random variables, the logarithm is in the basis of two.

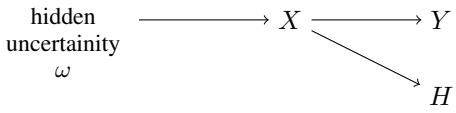

Fig. 1. Relationship between uncertain variables in non-stochastic hypothesis testing based on uncertain measurements.

B. Hypothesis Testing Based on Uncertain Measurements

Consider three uncertain variables in Figure 1. Uncertain variable X denotes an original uncertain variable. We have access to an *uncertain measurement* of this variable denoted by Y. This is captured by that $Y = g_Y(X)$ for a mapping $g_Y : [\![X]\!] \to [\![Y]\!]$. Recalling that uncertain variables are mappings from the sample space, it must be that $Y = g_Y \circ X$. Similarly, we may define the *hypothesis* as an uncertain variable H with binary range $[\![H]\!] = \{p_0, p_1\}$, where p_0 denotes the *null hypothesis* and p_1 denotes the *alternative hypothesis*. We assume that there exists a mapping $g_H : [\![X]\!] \to [\![H]\!]$ such that $H = g_H \circ X$; the hypothesis is constructed based on the uncertain variable X as $H = g_H(X)$. Note that Y and H are conditionally unrelated, conditioned on uncertain variable X.

A *test* is a function $T : [\![Y]\!] \to [\![H]\!] = \{p_0, p_1\}$. If $T(Y) = p_1$, the test rejects the null hypothesis in favour of the alternative hypothesis; however, if $T(Y) = p_0$, the test accepts the null hypothesis. The set of all tests is given by $[\![H]\!]^{[\![Y]\!]}$, which captures the set of all functions from $[\![Y]\!]$ to $[\![H]\!]$.

Consider $y \in [\![Y]\!]$ such that $T(y) = p_0$; the null hypothesis is accepted. The realization of output $Y(\omega) = y$ may correspond to many realizations of uncertain variable X, i.e., all the elements of the set $[\![X|y]\!]$. We say that $T(y) = p_0$ is correct, or the test is correct for the output realization $Y(\omega) = y$, if $g_H(x) = p_0$ for all $x \in [\![X|y]\!]$, i.e., all realizations of uncertain variable X compatible with y that are also compatible with the null hypothesis. The same holds for the alternative hypothesis as well. In following definition, we use the notation that $[\![H|[\![X|y]\!]]\!] := \{h \in [\![H|x]\!] : x \in [\![X|y]\!]\} = \cup_{x \in [\![X|y]\!]} [\![H|x]\!]$.

Definition 1 (Correctness): A test $T \in [\![H]\!]^{[\![Y]\!]}$ is correct at $y \in [\![Y]\!]$ if $[\![H|[\![X|y]\!]]\!] = \{T(y)\}$. The set of all outputs for which the test is correct is $\aleph(T) := \{y \in [\![Y]\!] : [\![H|[\![X|y]\!]]\!] = \{T(y)\}\}$. $\lhd$

Based on this definition of correctness, we can define a performance measure for a test:

$$\mathcal{P}(T) := \begin{cases} \log_e(\mu(\aleph)), & Y \text{ is a continuous u.v.} \\ \log_2(|\aleph|), & Y \text{ is a discrete u.v.} \end{cases} \tag{3}$$

We seek an optimal hypothesis test using the optimization problem in

$$T^* \in \underset{T \in [\![H]\!]^{[\![Y]\!]}}{\arg\max}\ \mathcal{P}(T). \tag{4}$$

Definition 2 (Consistency): A test $T : [\![Y]\!] \to [\![H]\!]$ is consistent if (*i*) $T(Y) = p_0$ only if $Y \in [\![Y|p_0]\!]$ and (*ii*) $T(Y) = p_1$ only if $Y \in [\![Y|p_1]\!]$. $\lhd$

In the following theorem, we prove that consistent tests are in fact optimal in the sense of $\mathcal{P}$. For any mapping $g : x \mapsto y$, we define the inverse image $g^{-1}(y) := \{x : g(x) = y\}$.

Theorem 1: (Optimal Tests): Any consistent test is a solution of (4). ◁

Proof: The proofs are removed due to page limits. See [42] for the complete proofs. ■

Note that if the realization of the lossy/uncertain measurement Y belongs to $[\![Y|p_0]\!] \cap [\![Y|p_1]\!]$, there is not enough evidence to accept or reject the null hypothesis or the alternative hypothesis. However, if the realization of the measurement Y belongs to $([\![Y|p_0]\!] \setminus [\![Y|p_1]\!]) \cup ([\![Y|p_1]\!] \setminus [\![Y|h_2]\!]) = [\![Y|p_0]\!] \Delta [\![Y|p_1]\!]$, with Δ denoting the symmetric difference operator on the sets, we can confidently reject or accept the null hypothesis or the alternative hypothesis. This fact is used by the consistent tests to achieve the highest performance.

Theorem 2: (Performance Bound): The performance of any test $T \in [\![H]\!]^{[\![Y]\!]}$ is upper bounded as

$$\mathcal{P}(T) \leq \begin{cases} \log_e(\mu([\![Y|p_0]\!] \Delta [\![Y|p_1]\!])), & Y \text{ is a continuous u.v.} \\ \log_2(|[\![Y|p_0]\!] \Delta [\![Y|p_1]\!]|), & Y \text{ is a discrete u.v.} \end{cases}$$

◁

Proof: The upper bound in the statement of the theorem follows from the proof of Theorem 1. ■

Theorem 2 can be seen as a non-stochastic equivalent of the Chernoff-Stein Lemma (see, e.g., [43, Ch. 11] for randomized hypothesis testing). Note that $\log_e(\mu([\![Y|p_0]\!] \Delta [\![Y|p_1]\!]))$ essentially captures the difference between the ranges $[\![Y|p_0]\!]$ and $[\![Y|p_1]\!]$ resembling the Kullback–Leibler divergence in a non-stochastic framework. A same interpretation can also be provided for $\log_2(|[\![Y|p_0]\!] \Delta [\![Y|p_1]\!]|)$.

Example 1: (Hypothesis Testing Using Noisy Measurements): Consider an uncertain variable $X = (X_1, X_2) \in [100, 250] \times [-10, 10]$, where X_1 denotes the height of an individual in centimetres and X_2 denotes the measurement error in centimetres. The uncertain measurement is $Y = g_Y(X) = X_1 + X_2$. Further, the hypothesis uncertain variable is defined as $H = g_H(X) = p_0 \mathbb{1}_{X_1 \leq 150} + p_1 \mathbb{1}_{X_1 > 150}$. The null hypothesis p_0 is that the individual's is short (i.e., shorter than or equal to 150 centimetres) and the alternative hypothesis is that the individual is tall (i.e., taller than 150 centimetres). Now, note that

$$\begin{aligned} [\![Y|p_0]\!] &= \{X_1 + X_2 : 100 \leq X_1 \leq 150, X_2 \in [-10, 10]\} \\ &= [90, 160], \\ [\![Y|p_1]\!] &= \{X_1 + X_2 : 150 \leq X_1 \leq 250, X_2 \in [-10, 10]\} \\ &= [140, 260]. \end{aligned}$$

Thus $[\![Y|p_0]\!] \cap [\![Y|p_1]\!] = [140, 160]$. Let T be a test such that $T(Y) = p_0$ if $Y \in [90, 150]$ and $T(Y) = p_1$ if $Y \in (150, 260]$. Evidently, T is a consistent test. We get

$$\begin{aligned} \mathcal{P}(T) &= \log_e(\mu([\![Y|p_0]\!] \Delta [\![Y|p_1]\!])) \\ &= \log_e(\mu([90, 140) \cup (160, 260])) \\ &= \log_e(150). \end{aligned}$$

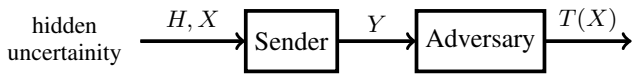

Fig. 2. Communication structure between a sender and a hypothesis-testing adversary.

Further, note that $h_0(Y) = \log_e(170)$. If we scale the performance by $h_0(Y)$, we get

$$\mathcal{P}(T) - h_0(Y) = \log_e(150) - \log_e(170) \approx -0.1251.$$

Now imagine the case where $X = (X_1, X_2) \in [100, 250] \times [-20, 20]$ with the interpretation that the amount of the additive uncertain measurement noise is twice larger. In this case, we have

$$\mathcal{P}(T) - h_0(Y) = \log_e(150) - \log_e(190) \approx -0.2364.$$

This shows that, by increasing the amount of the noise, the confidence of the test is reduced, which is in line with our expectation. ◁

III. Non-Stochastic Privacy Against Hypothesis-Testing Adversary

Consider the communication diagram in Figure 2 between a sender and an adversary. The adversary's ultimate aim is to accurately test a hypothesis H based on the communicated information from the sender Y. The sender wants to provide a message Y that is as close as possible to X while making the adversary's task in testing the validity of hypothesis H difficult. The policy of the sender is captured by the mapping from X to Y, denoted by g_Y. We use the performance of the adversary in testing the private hypothesis based on the reported output Y to define a measure of privacy as

$$\mathrm{Priv}(g_Y) := \begin{cases} h_0(Y) - \log_e(\mu([\![Y|p_0]\!] \Delta [\![Y|p_1]\!])), \\ \qquad\qquad Y \text{ is a continuous u.v.} \\ H_0(Y) - \log_2(|[\![Y|p_0]\!] \Delta [\![Y|p_1]\!]|)), \\ \qquad\qquad Y \text{ is a discrete u.v.} \end{cases} \tag{5}$$

Note that increasing $\mathrm{Priv}(g_Y)$ implies that the size of $[\![Y|p_0]\!] \Delta [\![Y|p_1]\!]$ is decreased, thus degrading the performance of any test employed by the adversary in light of Theorem 2.

Definition 3 (ϵ-privacy): Policy g_Y is ϵ-private for some $\epsilon \in (0, \infty)$ if $\mathrm{Priv}(g_Y) \geq \log(1 + \epsilon)$. ◁

We need to balance privacy with utility, otherwise the best policy is to report nothing (or to report at uncertain variable that is unrelated to X). Therefore, we need to define a measure of accuracy to balance against the privacy.

Definition 4 (ρ-accuracy): Policy g_Y is ρ-accurate for some $\rho \in (0, +\infty)$ if $\sup_{X \in [\![X]\!]} \|X - g_Y(X)\| \leq 1/\rho$. ◁

Increasing ρ in ρ-accuracy implies that $\sup_{X \in [\![X]\!]} \|X - Y\|$ is decreased, thus improving the quality of the reported output Y by enforcing it to stay consistently closer to X.

Theorem 3: Assume that $[\![X]\!] \subseteq \mathbb{R}^{n_x}$, and $g : \mathbb{R}^{n_x-1} \to \mathbb{R}$ exists such that

$$g_H(x) = \begin{cases} p_0, & x_i - g(x_{-i}) \geq 0, \\ p_1, & x_i - g(x_{-i}) < 0, \end{cases}$$

where $x_{-i} = (x_j)_{j \neq i}$. Let

$$g_Y(x) = \begin{cases} (g(x_{-i}), x_{-i}), & g(x_{-i}) - \dfrac{1}{\rho} \leq x_i \leq g(x_{-i}) + \dfrac{1}{\rho}, \\ x, & \text{otherwise}, \end{cases}$$

and

$$\epsilon = \left[\frac{\exp(h_0(X))}{\mu\left([\![X]\!] \cap \left\{ x : g(x_{-i}) - \dfrac{1}{\rho} \leq x_i \leq g(x_{-i}) + \dfrac{1}{\rho} \right\} \right)} - 1 \right]^{-1}.$$

Then, g_Y is ρ-accurate and ϵ-private. ◁

Proof: See [42]. ■

For large enough ρ, it can be seen that $[\![X]\!] \cap \{x : g(x_{-i}) - \rho^{-1} \leq x_i \leq g(x_{-i}) + \rho^{-1}\} \approx \{x : g(x_{-i}) - \rho^{-1} \leq x_i \leq g(x_{-i}) + \rho^{-1}\}$, and, as a result, $\epsilon = \mathcal{O}(\rho^{-1})$. This implies that, for the policy in Theorem 3, we have

"privacy × accuracy = constant".

With the theoretical results in hand, we can demonstrate the effects of using the policy in Theorem 3 on a practical dataset in the next section.

IV. Numerical Results

In this subsection, we consider design of a privacy-preserving policy for reporting individuals height in centimetres and weight in kilograms publicly. We consider an adversary who is interested in identifying individuals passing the obesity threshold in terms of body mass index (BMI), e.g., an insurance agency may use publicly available data to increase premiums of obese people or deny them insurance. Therefore, there is a duty of care when releasing demographic data of individuals publicly. By the definition of the U.S. Department of Health & Human Services, a person, be it female or male, is considered obese if their BMI is greater than or equal to 30.

Let uncertain variable X denote the weight and height, i.e., $X = [X_1 \;\; X_2]^\top$ with $X_1 \in [0, 200]$ denoting the weight in kilograms and $X_2 \in [0, 250]$ denoting the height in centimetres. The hypothesis is

$$g_H(x) = \begin{cases} p_0, & \dfrac{x_1}{(x_2/100)^2} \geq 30, \\ p_1, & \dfrac{x_1}{(x_2/100)^2} < 30. \end{cases}$$

Following the notation of Theorem 3, we can re-define the hypothesis using the sign of $x_1 - g(x_2)$ with $g : x_2 \mapsto 30(x_2/100)^2$. Define

$$g_Y(x) = \begin{cases} 30\left(\dfrac{x_2}{100}\right)^2, & \dfrac{3x_2^2}{1000} - \dfrac{1}{\rho} \leq x_1 \leq \dfrac{3x_2^2}{1000} + \dfrac{1}{\rho}, \\ x_1, & \text{otherwise}. \end{cases} \tag{6}$$

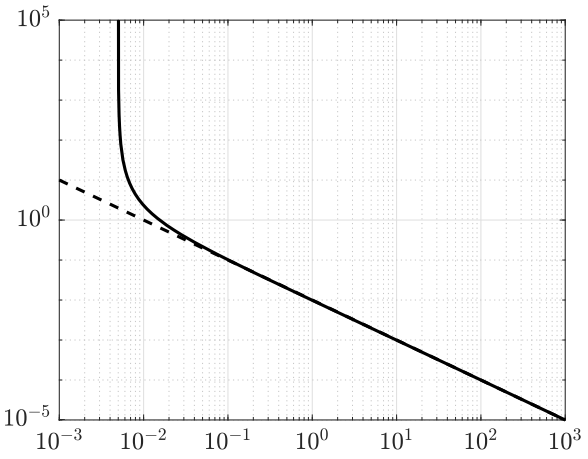

Fig. 3. Privacy guarantee, ϵ, versus accuracy level, ρ. The dashed line shows the asymptotic $\mathcal{O}(\rho^{-1})$.

Note that

$$\begin{aligned} \mu\left([\![X]\!] \cap \left\{ x : g(x_{-i}) - \frac{1}{\rho} \leq x_i \leq g(x_{-i}) + \frac{1}{\rho} \right\} \right) &= \int_0^{250} \int_{\max(3x_2^2/1000 - 1/\rho, 0)}^{\min(3x_2^2/1000 + 1/\rho, 200)} \mathrm{d}x_1 \mathrm{d}x_2 \\ &= \int_0^{250} [\min(3x_2^2/1000 + 1/\rho, 200) \\ &\qquad - \max(3x_2^2/1000 - 1/\rho, 0)] \mathrm{d}x_2. \end{aligned}$$

Using this, we can compute the level of privacy guarantee. The solid black line in Figure 3 illustrates privacy guarantee, ϵ, versus accuracy level, ρ. The dashed line shows the asymptotic $\mathcal{O}(\rho^{-1})$. As expected, by increasing accuracy, the privacy guarantee can only be reduced and *vice versa*.

Now, we use a real dataset to investigate the effects of the privacy-preserving policy in (6). We use a dataset of preferences, interests, and demographics of young people, aged between 15-30, of Slovakian nationality [44]. The data was gathered in 2013 by students of an statistics class at FSEV UK through friends and families. The dataset consists of 1,010 records with 150 features (139 integer and 11 categorical) including height, weight, music preferences, eating habits, etc. This dataset is popular for analysis on Kaggle (an online platform for sharing data) with 99.4k views and 21.8k downloads on all continents within the last three years. Noting that the preferences of the individuals can be matched with publicly available datasets, such as IMDb (Internet Movie Database), to potentially identify the individuals, there is a need for obfuscating the data in order to avoid privacy breaches related to age, weight, and height.

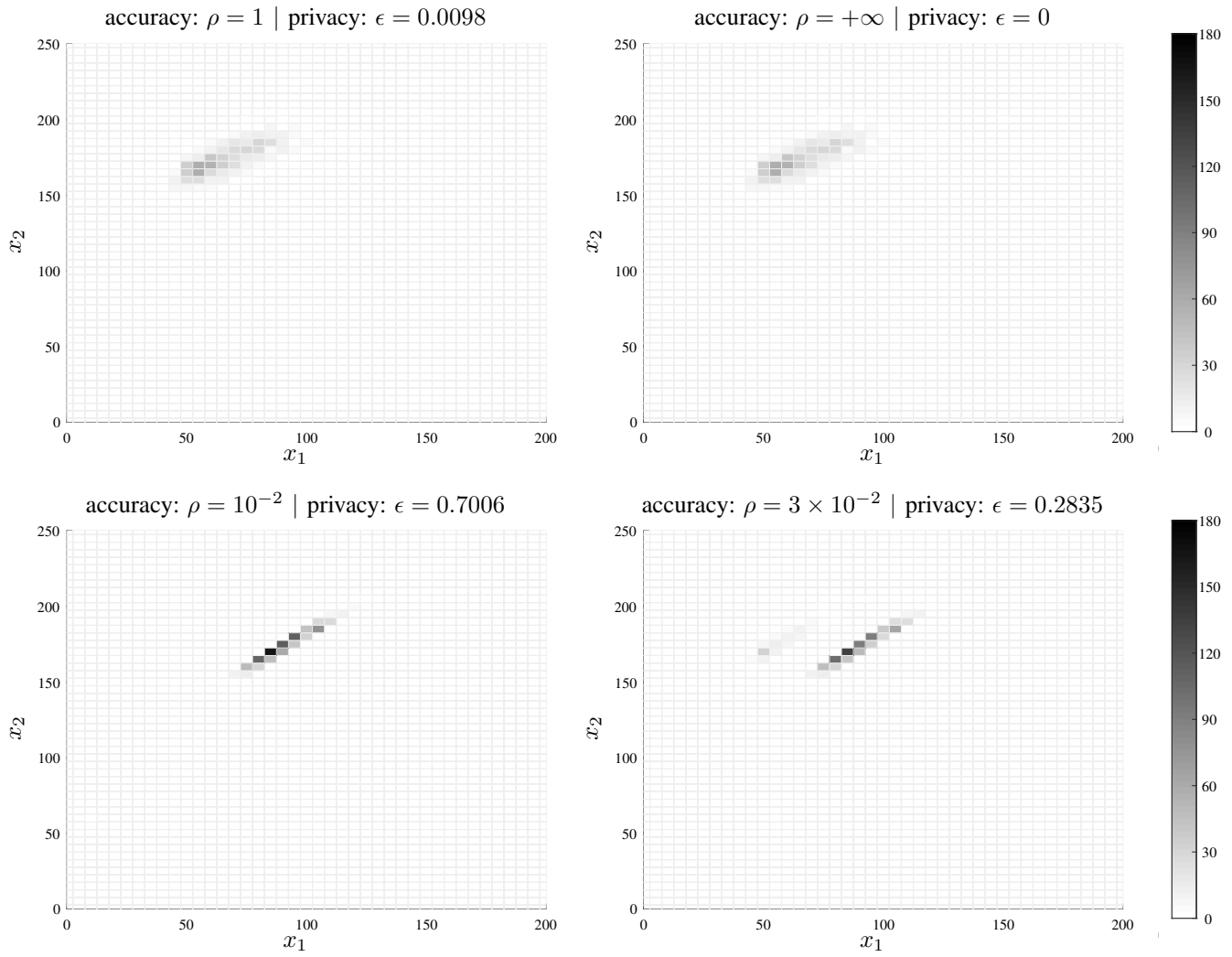

Fig. 4. The histogram of the reported weight and height of individuals for various levels of accuracy, ρ. The darker colors show higher frequencies.

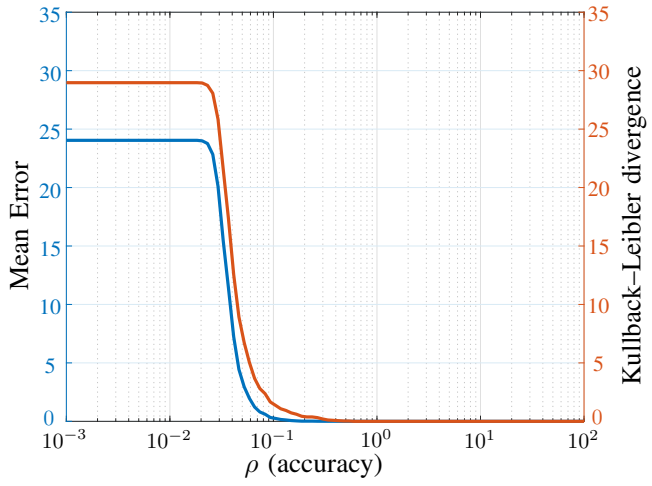

Fig. 5. The blue curve illustrates the difference of the mean value of the weight of the population with and without the privacy preserving policy. The red curve shows the difference between the empirical density functions of the weight across the population with and without the privacy preserving policy, measured by the Kullback–Leibler divergence.

Assume that we use the policy (6) is used for reporting weight and height of individuals so that potential future insurance agencies cannot test for the obesity levels. Figure 4 illustrates the histogram of the reported weight and hight of individuals for various levels of accuracy, ρ. The darker colors show higher frequencies. As expected, by decreasing ρ, the accuracy gets worse; the histogram changes more drastically. The blue curve in Figure 5 illustrates the difference of the mean value of the weight of the population with and without the privacy preserving policy. The red curve in Figure 5 shows the difference between the empirical density functions of the weight across the population with and without the privacy preserving policy. These curves allows a data curator to balance between privacy and utility.

V. Conclusions and Future Works

We considered privacy against hypothesis testing adversaries using the theory of non-stochastic hypothesis testing. We constructed reporting policies with prescribed privacy and utility guarantees. We demonstrated the utility of

privacy-preserving policies on a real dataset. Future work can focus on development of optimal policies.

REFERENCES

[1] S. L. Warner, "Randomized response: A survey technique for eliminating evasive answer bias," *Journal of the American Statistical Association*, vol. 60, no. 309, pp. 63–69, 1965.

[2] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Foundations and Trends in Theoretical Computer Science*, vol. 9, no. 3–4, pp. 211–407, 2014.

[3] J. C. Duchi, M. I. Jordan, and M. J. Wainwright, "Local privacy and statistical minimax rates," in *Foundations of Computer Science (FOCS), 2013 IEEE 54th Annual Symposium on*, pp. 429–438, IEEE, 2013.

[4] P. Kairouz, S. Oh, and P. Viswanath, "Extremal mechanisms for local differential privacy," in *Advances in Neural Information Processing Systems*, pp. 2879–2887, 2014.

[5] A. Machanavajjhala, D. Kifer, J. Abowd, J. Gehrke, and L. Vilhuber, "Privacy: Theory meets practice on the map," in *Proceedings of the 2008 IEEE 24th International Conference on Data Engineering*, pp. 277–286, IEEE Computer Society, 2008.

[6] N. Wu, F. Farokhi, D. Smith, and M. A. Kaafar, "The value of collaboration in convex machine learning with differential privacy," in *Security and Privacy, 2020. SP 2020. IEEE Symposium on*, IEEE, 2020.

[7] F. Farokhi and G. Nair, "Privacy-constrained communication," *IFAC-PapersOnLine*, vol. 49, no. 22, pp. 43–48, 2016.

[8] M. J. Wainwright, M. I. Jordan, and J. C. Duchi, "Privacy aware learning," in *Proceedings of Advances in Neural Information Processing Systems (NIPS)*, pp. 1430–1438, 2012.

[9] Y. Liang, H. V. Poor, and S. Shamai, "Information theoretic security," *Foundations and Trends in Communications and Information Theory*, vol. 5, no. 4–5, pp. 355–580, 2009.

[10] L. Lai, S.-W. Ho, and H. V. Poor, "Privacy–security trade-offs in biometric security systems–Part I: Single use case," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 1, pp. 122–139, 2011.

[11] F. Farokhi, H. Sandberg, I. Shames, and M. Cantoni, "Quadratic Gaussian privacy games," in *Proceedings of the 54th IEEE Conference on Decision and Control*, pp. 4505–4510, 2015.

[12] Z. Li and T. Oechtering, "Privacy on hypothesis testing in smart grids," in *IEEE Information Theory Workshop (ITW) 2015, Jeju, Korea, Oct. 11-15, 2015*, pp. 337–341, IEEE, 2015.

[13] F. Farokhi and H. Sandberg, "Fisher information as a measure of privacy: Preserving privacy of households with smart meters using batteries," *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 4726–4734, 2018.

[14] L. Sankar, S. R. Rajagopalan, and H. V. Poor, "Utility-privacy tradeoffs in databases: An information-theoretic approach," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 6, pp. 838–852, 2013.

[15] F. Farokhi and H. Sandberg, "Ensuring privacy with constrained additive noise by minimizing fisher information," *Automatica*, vol. 99, pp. 275–288, 2019.

[16] A. D. Wyner, "The wire-tap channel," *Bell System Technical Journal*, vol. 54, no. 8, pp. 1355–1387, 1975.

[17] T. Courtade, "Information masking and amplification: The source coding setting," in *Proceedings of the IEEE International Symposium on Information Theory Proceedings (ISIT)*, pp. 189–193, 2012.

[18] H. Yamamoto, "A source coding problem for sources with additional outputs to keep secret from the receiver or wiretappers," *IEEE Transactions on Information Theory*, vol. 29, no. 6, pp. 918–923, 1983.

[19] H. Yamamoto, "A rate-distortion problem for a communication system with a secondary decoder to be hindered," *IEEE Transactions on Information Theory*, vol. 34, no. 4, pp. 835–842, 1988.

[20] P. Samarati, "Protecting respondents identities in microdata release," *IEEE transactions on Knowledge and Data Engineering*, vol. 13, no. 6, pp. 1010–1027, 2001.

[21] L. Sweeney, "k-anonymity: A model for protecting privacy," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 10, no. 05, pp. 557–570, 2002.

[22] A. Machanavajjhala, J. Gehrke, D. Kifer, and M. Venkitasubramaniam, "ℓ-diversity: privacy beyond k-anonymity," in *22nd International Conference on Data Engineering (ICDE'06)*, pp. 24–24, 2006.

[23] A. Narayanan and V. Shmatikov, "Robust de-anonymization of large sparse datasets," in *Security and Privacy, 2008. SP 2008. IEEE Symposium on*, pp. 111–125, IEEE, 2008.

[24] J. Su, A. Shukla, S. Goel, and A. Narayanan, "De-anonymizing web browsing data with social networks," in *Proceedings of the 26th International Conference on World Wide Web*, pp. 1261–1269, 2017.

[25] Y.-A. De Montjoye, C. A. Hidalgo, M. Verleysen, and V. D. Blondel, "Unique in the crowd: The privacy bounds of human mobility," *Scientific reports*, vol. 3, p. 1376, 2013.

[26] F. Farokhi, "Development and analysis of deterministic privacy-preserving policies using non- stochastic information theory," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 10, pp. 2567–2576, 2019.

[27] J. Bambauer, K. Muralidhar, and R. Sarathy, "Fool's gold: an illustrated critique of differential privacy," *Vanderbilt Journal of Entertainment & Technology Law*, vol. 16, p. 701, 2013.

[28] F. Farokhi, J. Milosevic, and H. Sandberg, "Optimal state estimation with measurements corrupted by Laplace noise," in *Decision and Control (CDC), 2016 IEEE 55th Conference on*, pp. 302–307, IEEE, 2016.

[29] S. L. Garfinkel, J. M. Abowd, and S. Powazek, "Issues encountered deploying differential privacy," in *Proceedings of the 2018 Workshop on Privacy in the Electronic Society*, pp. 133–137, ACM, 2018.

[30] R. Bild, K. A. Kuhn, and F. Prasser, "SafePub: A truthful data anonymization algorithm with strong privacy guarantees," *Proceedings on Privacy Enhancing Technologies*, vol. 2018, no. 1, pp. 67–87, 2018.

[31] R. Bhaskar, A. Bhowmick, V. Goyal, S. Laxman, and A. Thakurta, "Noiseless database privacy," in *International Conference on the Theory and Application of Cryptology and Information Security*, pp. 215–232, Springer, 2011.

[32] S. U. Nabar, B. Marthi, K. Kenthapadi, N. Mishra, and R. Motwani, "Towards robustness in query auditing," in *Proceedings of the 32nd international conference on Very large data bases*, pp. 151–162, VLDB Endowment, 2006.

[33] R. V. L. Hartley, "Transmission of information," *Bell System Technical Journal*, vol. 7, no. 3, pp. 535–563, 1928.

[34] A. N. Kolmogorov and V. M. Tikhomirov, "ε-entropy and ε-capacity of sets in function spaces," *Uspekhi Matematicheskikh Nauk*, vol. 14, no. 2, pp. 3–86, 1959. English translation American Mathematical Society Translations, series 2, vol. 17, pp. 277–364.

[35] A. Renyi, "On measures of entropy and information," in *Proc. of the Fourth Berkeley Symp. on Math. Statist. and Prob.*, vol. 1, pp. 547–561, 1961.

[36] G. N. Nair, "A nonstochastic information theory for communication and state estimation," *IEEE Transactions on Automatic Control*, vol. 58, no. 6, pp. 1497–1510, 2013.

[37] D. Jagerman, "ε-entropy and approximation of bandlimited functions," *SIAM Journal on Applied Mathematics*, vol. 17, no. 2, pp. 362–377, 1969.

[38] G. N. Nair, "A nonstochastic information theory for feedback," in *Decision and Control (CDC), 2012 IEEE 51st Annual Conference on*, pp. 1343–1348, IEEE, 2012.

[39] P. Duan, F. Yang, S. L. Shah, and T. Chen, "Transfer zero-entropy and its application for capturing cause and effect relationship between variables," *IEEE Transactions on Control Systems Technology*, vol. 23, no. 3, pp. 855–867, 2015.

[40] M. Wiese, K. H. Johansson, T. J. Oechtering, P. Papadimitratos, H. Sandberg, and M. Skoglund, "Uncertain wiretap channels and secure estimation," in *Information Theory (ISIT), 2016 IEEE International Symposium on*, pp. 2004–2008, IEEE, 2016.

[41] R. F. Barber and J. Duchi, "Privacy: A few definitional aspects and consequences for minimax mean-squared error," in *53rd IEEE Conference on Decision and Control*, pp. 1365–1369, IEEE, 2014.

[42] F. Farokhi, "Non-stochastic hypothesis testing with application to privacy against hypothesis-testing adversary." Technical Note, Preprint: arXiv:1904.07377v1 [cs.IT], 2019. `https://arxiv.org/pdf/1904.07377v1.pdf`.

[43] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley, 2012.

[44] M. Sabo, "Young people survey: Explore the preferences, interests, habits, opinions, and fears of young people." available online at Kaggle.com, last visit: 8-Mar-2019. `https://www.kaggle.com/miroslavsabo/young-people-survey`.

Author/s:
Farokhi, F

Title:
Non-Stochastic Hypothesis Testing with Application to Privacy Against Hypothesis-Testing Adversaries

Date:
2020-03-12

Citation:
Farokhi, F. (2020). Non-Stochastic Hypothesis Testing with Application to Privacy Against Hypothesis-Testing Adversaries. Proceedings of the 2019 IEEE 58th Conference on Decision and Control (CDC), 2019-December, pp.6118-6123. IEEE. https://doi.org/10.1109/CDC40024.2019.9029652.